

사회과학 연구방법

: 질문지 관리와 분석 준비

5. 질문지 관리

• [참고] 자료수집방법별 특성

조사방법	우편조사 (Mail survey)	전화조사 (Telephone survey)	면접조사 (Face-to-Face Interview)
장점	1. 비용절감 2. 편리성 3. 응답시간의 비제약성 4. 응답의 신뢰성 5. 익명성 및 비밀보장 6. 조사원의 편견 감소 7. 광범위한 접근 가능	1. 신속한 자료 수집 2. 적은 비용 3. 익명성의 보장으로 심도깊은 질문이 가능 4. 광범위한 접근 가능 5. 응답의 연속성유지 가능	1. 융통성 2. 많은 질문과 복잡한 질문이 가능 3. 접근 곤란한 대상의 접촉이 가능 4. 높은 응답율 5. 지시사항의 응답가능
단점	1. 낮은 회수율 2. 긴 시간 소요 3. 무응답 항목 발생 4. 조사원의 기여 불가 5. 서술형 질문의 회피	1. 통제의 어려움 2. 신뢰성의 감소 3. 시청각 자료의 부족 4. 응답자의 제한 5. 적은 수의 질문	1. 비싼 비용 2. 조사원의 편견 가능 3. 응답자의 비협조적 4. 낮은 익명성 5. 안전성의 결여

1. 질문지 관리

1) 우편조사의 관리지침

- 전문적인 조사 설문지처럼 응답자가 다루기 편리한 소책자로 설계 : 자기기입식 특성 고려
- 인사말과 협조에 감사하는 말을 표지와 책자의 마지막 부분에 작성
- 설문지 회송을 위한 우편 주소 및 안내문 포함(반송용 회송봉투/수취인부담 제공)
 - . 회수 상황 모니터링 및 회수율 관리
- 설문지 고유번호 작성, 응답 내용에 대한 비밀보장 안내문 삽입
- 문의 전화번호 및 담당자(연구기관 및 조사기관) 제공

* 확인우편

- 처음 설문지를 발송한 날로부터 2~3주 후에 확인 우편 재발송
- 특별한 표본추출방법에 채택
- 8~15%정도의 회수율이 기대되므로 7~12배수 정도의 표본 선정 : 최근 대표성 문제 제기

2) 전화조사 관리지침

* 전화조사의 성공의 관건

- 적당한 전화 조사원의 선정
- 전화 조사원의 훈련 : 특정 조사항목을 다루는 방법 훈련

* 전화조사 면접요령

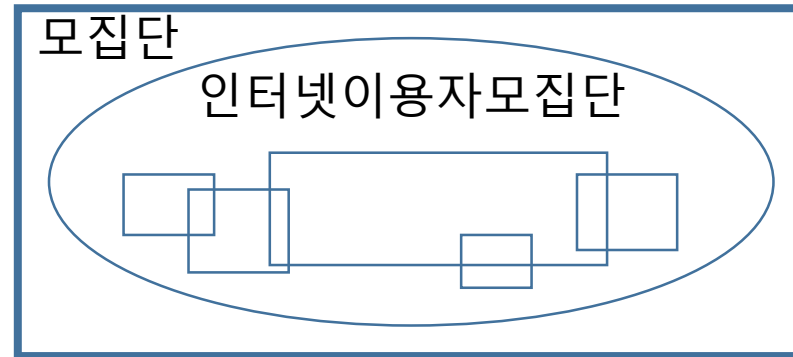
- 같은 가구에 대해 두 사람 이상의 성인을 조사하지 말 것
- 조사원의 친구나 친지를 대상으로부터 조사하지 말 것
- 다른 방해요소가 없도록 비밀리에 조사
- 조사를 타인에게 의뢰하지 말 것
- 조사를 거짓으로 행하지 말 것

* CATI system

- RDD
- 전문 조사원
- 녹음
- 모니터링

- [참고] Internet(or Web) survey

- 모집단 구조



- 인터넷조사 특성

- 응답 질 우수 : high quality response
 - Sampling : coverage, 대표성(representativeness)
 - Measurement : mode effect (the panel propensity, low response rate etc), 편향(bias)

- 현실

- 자료수집을 위한 보완적인 도구로 널리 활용 : not sampling, but collecting data
 - 대표성 한계
 - 추출틀 영향
 - 주로 마케팅조사에서 널리 활용

6. 분석을 위한 준비

1. 질문 문항에 대한 사전 코드화(precoding)

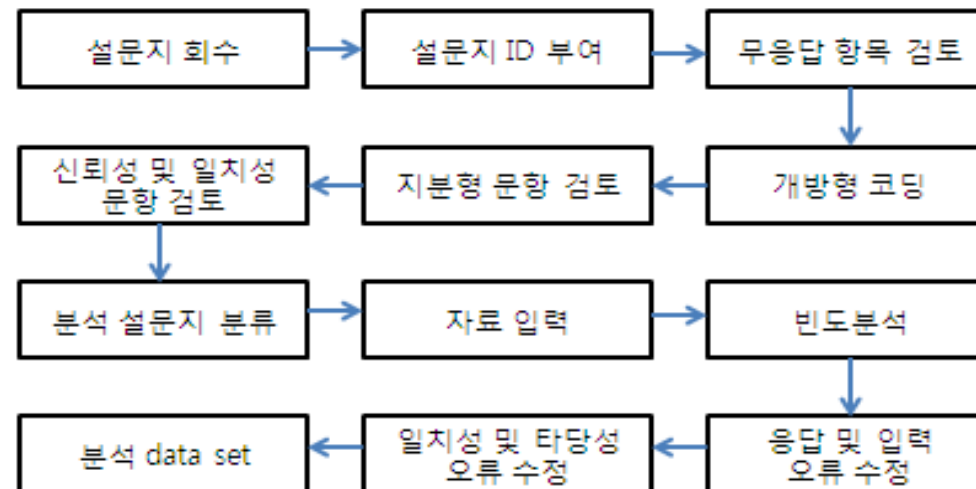
- 1) 사전코드 : 미리 설문지의 문항 또는 응답 항목에 코드를 부여하는 것
- 2) 코드(code) : 자료 입력 및 분석을 위해 문항 번호, 형식 및 입력 크기 등의 일관성 있도록 설정한 자료 입력 형식
- 3) 선택형 질문의 응답항목만 사전코드 가능
- 4) 입력공간 확보
- 5) 예비코드를 준비하여 입력시 오차 확인 및 수정을 위해 조사원코드와 응답자코드를 부여
- 6) 최근 조사 상황에 대한 자료(응답 시간, 조사 관심도, 조사에 대한 반응 등)를 분석 및 품질 제고를 위한 자료로 활용하기 위한 조사 관리 자료를 구축하기 위해 입력 자료를 미리 선정

2. 데이터 editing과 cleaning 과정

- 목적

- . 응답 및 무응답 확인
- . 조사 오차 감소 : 비표본오차
- . 유효 표본 확보
- . 품질 향상
 - + 정확성
 - + 신뢰성

분석 자료 클리닝 과정



3. 무응답의 처리

1) 무응답 요인

- 부재
- 거부
- 표본 부적격 등

2) 무응답의 영향

- 응답 분포 : 응답자와 무응답자의 분포
- 추정 및 추론 결과 왜곡
- 가중치 산출 영향
- 분산 증가

3) 무응답의 유형

- 단위무응답
- 항목무응답

4) 무응답 검토 및 수정

(1) 응답해야 할 모든 문항의 응답 여부를 확인

- 단위무응답(unit nonresponse) :
 - (1) 표본 비접촉/부재 : 표본 적격 미지
 - (2) 부적격 표본 : 표집틀의 포함 오류
 - (3) 조사 참여 거부 : 교체 가능
- 항목무응답(item nonresponse)

(2) 무응답 문항에 대한 처리 및 수정

- 기본적으로 재조사가 원칙
- 무응답 처리 방안
 - ① 그대로 두는 경우
 - 자료 입력 시 9, 99, 999 등 입력 자리 숫자의 최대값으로 부여
 - 무응답 코딩은 모수 추정 시 제외
 - ② 대체(imputation)
 - 평균대체
 - 회귀대체
 - Deck : Cold deck, Hot deck
 - 다중대체
 - 주의 : 대체 자료는 대체여부를 하나의 변수로 처리해 분산 추정 시 제외

[참고] 표본조사의 오차

$$\begin{array}{ccccc} \text{오 차} & \Rightarrow & \text{표본오차} & + & \text{비표본오차} \\ (\text{Error}) & & (\text{Sampling Error}) & & (\text{Nonsampling Error}) \end{array}$$

- 비표본오차 ; 실제 조사 과정이나 자료의 입력, 분석과 같은 자료의 처리과정에서 발생하는 오차로써 표본오차 이외의 모든 오차를 의미(systematically error)하며, 모든 조사에서 항상 발생되며 기본적으로는 통제가 불가능한 오차
- 비표본오차의 발생 원인
 - 1) 실제 조사 과정
 - . (단위) 무응답(nonresponse) : 표본 누락, 부재, 응답불능, 조사불능
 - . 부정확한 응답(inaccurate response)
 - . 표본선택의 편의(selection bias)
 - . 조사원의 편의(interviewer bias)
 - 2) 자료 처리 및 분석 과정
 - . 잘못 측정하거나 잘못 기재(기록)하는 경우
 - . 측정도구의 오류
 - . 자료입력 시 오류(punching error)
 - . 분석방법의 오류

● 비표본오차의 최소화 방법 ;

- 표준화된 설문지의 이용과 조사원의 훈련 및 감독, 조사 자료의 확인/검증 등 실제 조사 과정과 자료입력, 분석 등 자료처리 과정에서 매우 세심하게 주의를 기울이면 어느 정도 통제가 가능하여 오차를 감소시킬 수 있음

- 1) 재조사(Callback)
- 2) 응답에 대한 보상(reward)이나 동기유발(incentive)
- 3) 잘 훈련된 조사원 고용(trained interviewer)
- 4) 자료의 검증(data check)
- 5) 설문지 작성의 표준화

● 비표본오차의 유형과 해결 방안

- 무응답 : 재조사
 - 단위무응답 : 가중치 조정
 - 항목 무응답 : 무응답 대체
- 추출틀에 의한 오차 : 벤치마킹/사후층화 가중치
- 측정오차 : 표준화 측정도구 사용, 재조사(검증) 등
- 조사원에 의한 오차 : 검증(에디팅)
- 자료 입력 및 처리, 분석에 의한 오차 : 검증(에디팅)

4. 코딩

• 개방형 코딩

1) 코딩(coding)이란? 응답자가 직접 기록한 응답에 숫자나 부호를 부여하는 과정

1) 예 : 남자는 1, 여자는 0

2) 색깔 : 노랑 1, 빨강 2, 청색 3, 하늘색 4, 검정 5 등

2) 선택형에서의 코딩

- 선택형으로 부여된 숫자 이후의 번호를 순서대로 부여

C2. 사용(복용)하지 않은 약을 보관하는 이유는 무엇입니까?

① 나중에 사용하기 위해서

② 버리는 것을 잊어버려서

③ 무엇을 할지 잘 모르겠다.

④ 버릴려고 했었다.

⑤ 기타_____

3) 개방형에서의 코딩

- 설문지를 보면서 입력 번호를 부여

- 여러 명이 코딩작업을 할 때 통일된 번호로 코딩하기 위한 관리가 요구됨

D7. 보유하고 계신 만성질환은 무엇입니까?

- 지분형 문항의 검토 및 수정

예)

문22 현재 재택전자(인터넷, PC통신) 민원처리제가 시행되고 있다는 사실을 알고 계십니까?

① 예 ② 아니요 → 문 23으로 가시오

문22-1. 그러면 재택전자 민원처리제에 대해 어느 정도 만족하십니까?

전혀 불만족	대체로 불만족	보통	대체로 만족	매우 만족	잘모름 /무응답
①	②	③	④	⑤	<input type="checkbox"/> ⑥

- 지분을 위한 문항의 응답과 지분형 문항의 대상자가 일치하는지 검토
 - 지분 문항의 응답 기준으로 처리
 - 참고 : 빈도분석으로 검토하는 경우
 - 지분 기준 문항 & 지분형 무응답/결측 확인
 - 지분형 응답 여부 & 지분 기준 문항 확인
- 무응답과 결측(missing value)
 - 무응답 : 응답을 해야 하는데 응답하지 않은 것으로 입력 자료 부여(9, 99 등)
 - 결측 : 응답 대상이 아니므로 응답하지 않은 것으로 입력 자료 없음(결측 처리)

5. 신뢰성 및 일치성 문항 검토 및 수정

1) 신뢰성 문항 검토

- 응답자가 문항에 거짓없이 응답(측정오류 확인)하는지 확인하기 위해 작성된 문항으로 설문지내 서로 유사한 문항을 작성해 확인
- 유사 문항에 대한 응답이 다른 경우, 응답자가 설문 문항에 전반에 대해 거짓 혹은 편향되게 응답한 것으로 판단해 분석자료에서 제외하는 것이 바람직

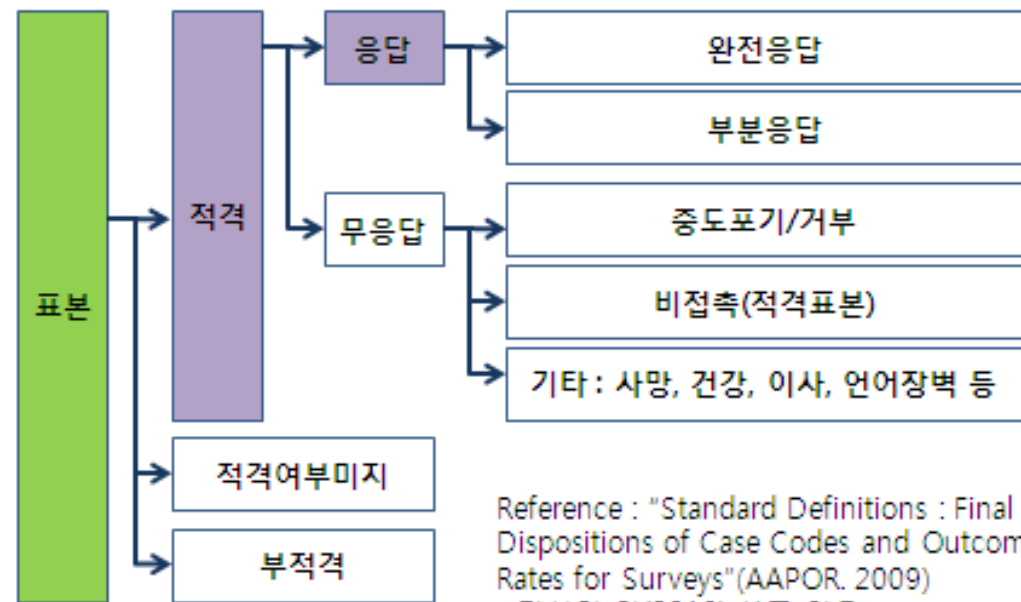
2) 일치성 문항 검토 및 수정

- 응답자가 수치 자료로 응답하는 경우, 수치 자료의 일치성을 검토해 수정하는 과정
- 수정 기준 : 응답자의 응답자료를 기준으로 판단하는 것이 바람직
 - (1) 응답 자료를 기준으로 합을 수정하는 방법
 - (2) 합을 기준으로 응답 자료를 수정하는 방법

6. 분석 설문지 분류/유효 분석 표본의 결정

- 표본의 접촉 및 응답 여부에 따른 분류

표본의 응답 결과 유형



Reference : "Standard Definitions : Final Dispositions of Case Codes and Outcome Rates for Surveys"(AAPOR, 2009)

- 김서영 외(2010) 참조 인용

1) 분석 제외 자료의 분류 기준

- [참고] **응답에 대한 표준 정의** (AAPOR, 2009) : 전화, 면접, 우편조사 적용

- ① 반드시 응답해야 할 문항 중 50% 미만(중도포기), 50~80% 미만(부분응답), 80% 이상 (완전응답)으로 간주
- ② 전체 문항 중 50% 미만(중도포기), 50~80% 미만(부분응답), 80% 이상 (완전응답)으로 간주
- ③ 핵심 질문 중 응답한 항목이 전체 핵심 질문의 50% 미만(중도포기), 50~99% 이하 (부분응답), 100% (완전응답)으로 간주
- ④ 위의 세 가지를 조합하여 정의 가능 : 핵심질문은 완전응답, 그 외 질문은 80% 이상 응답만을 완전응답으로 간주

2) 응답 제외 설문지 기준

- (1) 일반적으로 응답 문항 전체 중 90% 미만인 설문지 : 연구자의 기준 마련이 필요
- (2) 분석을 위한 기본 문항은 모두 응답하고 나머지 문항 중 80% 미만인 설문지
- (3) 신뢰성 문항에서 신뢰성에 의심이 가는 설문지

3) 응답 제외 설문지의 처리

- 분석 제외 자료는 단위 무응답 자료로 간주해 무응답 조정 가중치 부여
- 보고서 작성 시 분석 제외 자료 현황을 제시
 - (1) 전체 수집 자료 크기
 - (2) 분석 제외 자료 크기 : 제외 사유 표시
 - (3) 유효 분석 자료 크기

7. 자료 입력

1) 문항별로 변수명을 부여

- 변수명은 분석하는 통계 패키지의 변수 이름 부여 규칙에 따라 부여
- 데이터셋에 변수를 입력할 때 하나의 답은 하나의 변수로 입력하는 것이 원칙
 - 하나의 변수에 답에 2개 존재할 수 없음
- 예 : [문항 27] 다음 중에서 좋아하는 두 가지를 선택하시오.
 - 첫 번째 답 : q27_1
 - 두 번째 답 : q27_2

2) 입력 방법

- (1) 간접 입력 방법 : 엑셀, 메모장, 한글 등
- (2) 직접 입력 방법 : 분석 프로그램의 자료 입력 창에서 입력

8. 빈도분석을 이용한 cleaning

1) 자료 클리닝에서의 빈도분석

- 목적 :

- (1) 응답 및 입력 자료의 오류 수정을 위한 기초 분석
- (2) 수치 자료의 일치성 검토 및 확인을 위한 기초 분석
- (3) 연속형 자료의 범주화를 위한 기초 분석

2) 응답 및 입력 오류 자료의 검토 및 수정

- 모든 문항에 대한 빈도분석을 수행
- 문항별로 응답 범주를 벗어난 이상값 검토 및 수정

3) 일치성 검토 및 수정

- 수치 자료의 일치성을 확인하기 위해 개별 자료의 합과 응답자료 합의 일치 여부를 확인
- 수정 과정은 입력 전 일치성 검토 및 수정 과정과 동일

4) 연속형 수치 자료에 대한 범주화 변환

- 범주화 변환 방법
 - (1) 연속형 수치 자료의 범주화 작업은 통계적 이론을 근거로 범주화 시도
 - (2) 사회적 관행에 따른 범주화
- 범주화 변환 시 주의 사항
 - (1) 하나의 값이 두 개 범주에 속하지 않도록 주의 :
 - 미만, 이하/이상 사용 주의
 - (2) 범주화한 변수명을 원자료의 변수와 구분되게 부여

9. 자료분석에 대한 COMMENTS

1) 1차원 분석

(1) 빈도분석

(2) 기술통계량 분석 : 무응답 자료 제외함에 주의

2) 2차원 분석 : 응답자 특성과 관심 문항, 관심 문항과 관심 문항 사이의 2차원 분석

(1) 분석 전 검토 사항

- 양적 자료는 정규성(필요시 등분산성, 독립성) 확인이 기본
- 정규성 위배시 비모수 검정방법으로 분석

(2) 분석 방법

- 질적+질적 자료 : 카이제곱 검정
- 질적+양적 자료 : t검정, ANOVA
- 양적+양적 자료 : 상관분석

3) 인과관계 분석 : 다변량분석 포함

4) 분석 시 주의 사항

(1) 확률표본추출에 의한 표본조사는 반드시 가중치를 이용하여 추정과 분석이 필요

- 가중치(weight) : 표본단위에 대한 모집단 확대계수
- 가중치를 반영한 전문분석소프트웨어 사용
- 가중치 효과 :
 - 편향 제거, 분산 증가
 - 가중치 미반영시 편향, 분산 과소 추정의 문제 대두

(2) 확률표본조사 자료는 표본오차 관련 정보 제공